

Sobolev seminorm of quadratic functions with applications to derivative-free optimization

Zaikun Zhang

Received: date / Accepted: date

Abstract This paper studies the H^1 Sobolev seminorm of quadratic functions. The research is motivated by the least-norm interpolation that is widely used in derivative-free optimization. We express the H^1 seminorm of a quadratic function explicitly in terms of the Hessian and the gradient when the underlying domain is a ball. The seminorm gives new insights into least-norm interpolation. It clarifies the analytical and geometrical meaning of the objective function in least-norm interpolation. We employ the seminorm to study the extended symmetric Broyden update proposed by Powell. Numerical results show that the new theory helps improve the performance of the update. Apart from the theoretical results, we propose a new method of comparing derivative-free solvers, which is more convincing than merely counting the numbers of function evaluations.

Keywords Sobolev seminorm · Least-norm interpolation · Derivative-free optimization · Extended symmetric Broyden update

Mathematics Subject Classification (2010) 90C56 · 90C30 · 65K05

1 Motivation and Introduction

Consider an unconstrained derivative-free optimization problem

$$\min_{x \in \mathbb{R}^n} F(x), \quad (1.1)$$

where F is a real-valued function whose derivatives are unavailable. Problems of this type have numerous applications. For instance, they have been employed to solve the helicopter rotor blade design problem [4, 3, 2], the ground water community problems [27, 25], and the problems in biomedical imaging [36, 37].

Many algorithms have been developed for problem (1.1). For example, direct search methods (see Wright [53], Powell [40], Lewis, Torczon, and Trosset [33], and Kolda, Lewis,

Partially supported by Chinese NSF grants 10831006, 11021101, and CAS grant kjcx-yw-s7.

Zaikun Zhang
Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100190, CHINA
E-mail: zhangzk@lsec.cc.ac.cn

and Torczon [31] for reviews), line search methods without derivatives (see Stewart [48] for a quasi-Newton method using finite difference; see Gilmore and Kelley [26], Choi and Kelley [8], and Kelley [29, 30] for Implicit Filtering method, a hybrid of quasi-Newton and grid-based search; see Diniz-Ehrhardt, Martínez, and Raydán [22] for a derivative-free line search technique), and model-based methods (for instance, a method by Winfield [52], COBYLA by Powell [39], DFO by Conn, Scheinberg, and Toint [12, 13, 14], UOBYQA by Powell [41], a wedge trust region method by Marazzi and Nocedal [34], NEWUOA and extensions by Powell [44, 45, 46, 47], CONDOR by Vanden Berghen and Bersini [49], BOOSTERS by Ouevray and Bierlaire [38], ORBIT by Wild, Regis, and Shoemaker [51], and MNH by Wild [50]). We refer the readers to the book by Brent [6], the one by Kelley [28], and the one by Conn, Scheinberg, and Vicente [15] for extensive discussions and the references therein.

Multivariate interpolation has been acting as a powerful tool in the design of derivative-free optimization methods [52, 39, 16, 12, 13, 14, 5, 41, 34, 42, 44, 20, 19, 49, 17, 45, 51, 50, 38, 46, 47, 18]. The following quadratic interpolation plays an important role in Conn and Toint [16], Conn, Scheinberg, and Toint [12, 13, 14], Powell [43, 44, 45, 46, 47], and Custódio, Rocha, and Vicente [18]:

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla^2 Q\|_F^2 + \sigma \|\nabla Q(x_0)\|_2^2 \\ \text{s.t.} \quad & Q(x) = f(x), \quad x \in \mathcal{S}, \end{aligned} \tag{1.2}$$

where \mathcal{Q} is the linear space of polynomials of degree at most two in n variables, x_0 is a specific point in \mathbb{R}^n , σ is a nonnegative constant, \mathcal{S} is a finite set of interpolation points in \mathbb{R}^n , and f is a function¹ on \mathcal{S} . We henceforth refer to this type of interpolation as least-norm interpolation.

The objective function of problem (1.2) is interesting. It is easy to handle, and practice has proved it successful. The main purpose of this paper is to provide a new interpretation of it. Our tool is the H^1 Sobolev seminorm, which is classical in PDE theory but may be rarely noticed in nonlinear optimization. We will give new insights into some basic questions about the objective function in (1.2). For example, *what is the exact analytical and geometrical meaning of this objective?* Why to combine the Frobenius norm of the Hessian and the 2-norm of the gradient? What is the meaning of the parameters x_0 and σ ?

This paper is organized as follows. Section 2 introduces the applications of least-norm interpolation in derivative-free optimization. Section 3 presents the main theoretical results. We first study the H^1 seminorm of quadratic functions, and then employ the H^1 seminorm to investigate least-norm interpolation. The questions listed above are answered in Section 3. Section 4 applies the theory obtained in Section 3 to study the extended symmetric Broyden update proposed by Powell [47]. We show an easy and effective way to improve the performance of the update. Section 5 concludes our discussion.

The main contributions of the paper lie in Section 3 and Section 4. Besides the theoretical results, the numerical experience in Section 4 is also a highlight. We show an interesting numerical example which suggests that it is inadequate to compare derivative-free solvers by simply counting the numbers of function evaluations. To obtain more convincing comparison, we propose a new method of testing derivative-free solvers by introducing statistics into the numerical experiments. See Subsection 4.3 for details.

¹ Notice that f is not always equal to the objective function F , which is the case in Powell [43, 44, 45, 46, 47]. See Section 2 for details.

A remark on notation. We use the notation

$$\begin{aligned} & \min_{x \in \mathcal{X}} \psi(x) \\ & \text{s.t. } \min_{x \in \mathcal{X}} \phi(x) \end{aligned} \quad (1.3)$$

for the bilevel programming problem

$$\begin{aligned} & \min \psi(x) \\ & \text{s.t. } x \in [\arg \min_{\xi \in \mathcal{X}} \phi(\xi)]. \end{aligned} \quad (1.4)$$

2 Least-Norm Interpolation in Derivative-Free Optimization

Least-norm interpolation has successful applications in derivative-free optimization, especially in model-based algorithms. These algorithms typically follow trust-region methodology [11, 42]. On each iteration, a local model of the objective function is constructed, and then it is minimized within a trust-region to generate a trial step. The model is usually a quadratic polynomial constructed by solving an interpolation problem

$$Q(x) = F(x), \quad x \in \mathcal{S}, \quad (2.1)$$

where, as stated in problem (1.1), F is the objective function, and \mathcal{S} is an interpolation set. To determine a unique quadratic polynomial by problem (2.1), the size of \mathcal{S} needs to be at least $(n+1)(n+2)/2$, which is prohibitive when n is big. Thus we need to consider underdetermined quadratic interpolation. In that case, a classical strategy to take up the remaining freedom is to minimize some functional subject to the interpolation constraints, that is to solve

$$\begin{aligned} & \min_{Q \in \mathcal{Q}} \mathcal{F}(Q) \\ & \text{s.t. } Q(x) = F(x), \quad x \in \mathcal{S}, \end{aligned} \quad (2.2)$$

\mathcal{F} being some functional on \mathcal{Q} . Several existing choices of \mathcal{F} lead to least-norm interpolation, directly or indirectly. Here we give some examples.

Conn and Toint [16] suggests the quadratic model that solves

$$\begin{aligned} & \min_{Q \in \mathcal{Q}} \|\nabla^2 Q\|_F^2 + \|\nabla Q(x_0)\|_2^2 \\ & \text{s.t. } Q(x) = F(x), \quad x \in \mathcal{S}, \end{aligned} \quad (2.3)$$

where x_0 is a specific point in \mathbb{R}^n . Problem (2.3) is a least-norm interpolation problem with $\sigma = 1$. It is reported that the resultant algorithm is substantially more robust than a code based on LANCELOT [10] subroutines and finite-differences, although no motivation for (2.3) is given.

Conn, Scheinberg, and Toint [12, 13, 14] builds a quadratic model by solving

$$\begin{aligned} & \min_{Q \in \mathcal{Q}} \|\nabla^2 Q\|_F^2 \\ & \text{s.t. } Q(x) = F(x), \quad x \in \mathcal{S}. \end{aligned} \quad (2.4)$$

It is considered as the best general strategy for the DFO algorithm to select a unique model from the pool of the possible interpolation models [14]. Wild [50] also works with the model defined by (2.4). Problem (2.4) is a least-norm interpolation problem with $\sigma = 0$. Conn, Scheinberg, and Vicente [15] explains the motivation for (2.4) by the following error bounds of quadratic interpolants.

Theorem 2.1 [15] *Suppose that the interpolation set $S = \{y_0, y_1, \dots, y_m\}$ ($m \geq n$) is contained in a ball $\mathcal{B}(y_0, r)$ ($r > 0$), and the matrix*

$$L = \frac{1}{r}(y_1 - y_0 \cdots y_m - y_0) \quad (2.5)$$

has rank n . If F is continuously differentiable on $\mathcal{B}(y_0, r)$, and ∇F is Lipschitz continuous on $\mathcal{B}(y_0, r)$ with constant $\nu > 0$, then for any quadratic function Q satisfying the interpolation constraints (2.1), it holds that

$$\|\nabla Q(x) - \nabla F(x)\|_2 \leq \frac{5\sqrt{m}}{2} \|L^\dagger\|_2 (\nu + \|\nabla^2 Q\|_2) r, \quad \forall x \in \mathcal{B}(y_0, r), \quad (2.6)$$

and

$$|Q(x) - F(x)| \leq \left(\frac{5\sqrt{m}}{2} \|L^\dagger\|_2 + \frac{1}{2} \right) (\nu + \|\nabla^2 Q\|_2) r^2, \quad \forall x \in \mathcal{B}(y_0, r). \quad (2.7)$$

In light of Theorem 2.1, minimizing some norm of $\nabla^2 Q$ will help improve the approximation of gradient and function value. Notice that $\nabla^2 Q$ appears in (2.6–2.7) with 2-norm rather than Frobenius norm.

Powell [43, 44, 45] introduce the symmetric Broyden update to derivative-free optimization, and it is proposed to solve

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 \\ \text{s.t.} \quad & Q(x) = F(x), \quad x \in S \end{aligned} \quad (2.8)$$

to obtain a model for the current iteration, provided that Q_0 is the quadratic model used in the previous iteration. Let $D = Q - Q_0$ and $f = F - Q_0$, and then (2.8) is equivalent to

$$\begin{aligned} \min_{D \in \mathcal{Q}} \quad & \|\nabla^2 D\|_F^2 \\ \text{s.t.} \quad & D(x) = f(x), \quad x \in S. \end{aligned} \quad (2.9)$$

Thus problem (2.8) is essentially a least-norm interpolation problem about $Q - Q_0$. The symmetric Broyden update is motivated by least change secant updates [21] in quasi-Newton methods. One particularly interesting advantage of the update is that, when F is a quadratic function, the solution Q_+ of problem (2.8) has the property [43, 44, 45] that

$$\begin{aligned} & \|\nabla^2 Q_+ - \nabla^2 F\|_F^2 \\ &= \|\nabla^2 Q_0 - \nabla^2 F\|_F^2 - \|\nabla^2 Q_+ - \nabla^2 Q_0\|_F^2 \\ &\leq \|\nabla^2 Q_0 - \nabla^2 F\|_F^2. \end{aligned} \quad (2.10)$$

Thus $\nabla^2 Q_+$ approximates $\nabla^2 F$ better than $\nabla^2 Q_0$ unless $\nabla^2 Q_+ = \nabla^2 Q_0$. Additionally, if (2.8) is employed on every iteration, then (2.10) implies that the difference $\nabla^2 Q_+ - \nabla^2 Q_0$ will eventually converge to zero, which helps local convergence, in light of the classical theory in Broyden, Dennis, and Moré [7].

Powell [47] proposes the extended symmetric Broyden update by adding a first-order term to the objective function of problem (2.8), resulting in

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 + \sigma \|\nabla Q(x_0) - \nabla Q_0(x_0)\|_2^2 \\ \text{s.t.} \quad & Q(x) = F(x), \quad x \in \mathcal{S}, \end{aligned} \quad (2.11)$$

x_0 and σ being specifically selected parameters, σ nonnegative. Again, (2.11) can be interpreted as a least-norm interpolation about $Q - Q_0$. Powell [47] motivates (2.11) by an algebraic example for which the symmetric Broyden update does not behave satisfactorily. We will study the extended symmetric Broyden update in Section 4.

Apart from model-based methods, least-norm interpolation also has applications in direct search methods. Custódio, Rocha, and Vicente [18] incorporates models defined by (2.4) and (2.8) into direct search. The authors attempt to enhance the performance of direct search methods by taking search steps based on these models. It is reported that (2.4) works better for their method, and their procedure provides significant improvements to direct search methods of directional type.

For more discussions about least-norm interpolation in derivative-free optimization, we refer the readers to Chapter 5 and 11 of Conn, Scheinberg, and Vicente [15].

3 The H^1 Sobolev Seminorm of Quadratic Functions

In Sobolev space theory [1, 24], the H^1 seminorm of a function f over a domain Ω is defined as

$$|f|_{H^1(\Omega)} = \left[\int_{\Omega} \|\nabla f(x)\|_2^2 dx \right]^{1/2}. \quad (3.1)$$

In this section, we give an explicit formula for the H^1 seminorm of quadratic functions when Ω is a ball, and accordingly present a new understanding of least-norm interpolation. We prove that least-norm interpolation essentially seeks the quadratic interpolant with minimal H^1 seminorm over a ball. Theorem 3.1, Theorem 3.2, and Theorem 3.3 are the main theoretical results of this paper.

3.1 The H^1 Seminorm of Quadratic Functions

The H^1 seminorm of a quadratic function over a ball can be expressed explicitly in terms of its coefficients. We present the formula in the following theorem.

Theorem 3.1 *Let x_0 be a point in \mathbb{R}^n , r be a positive number, and*

$$\mathcal{B} = \{x \in \mathbb{R}^n : \|x - x_0\|_2 \leq r\}. \quad (3.2)$$

Then for any $Q \in \mathcal{Q}$,

$$|Q|_{H^1(\mathcal{B})}^2 = V_n r^n \left[\frac{r^2}{n+2} \|\nabla^2 Q\|_F^2 + \|\nabla Q(x_0)\|_2^2 \right], \quad (3.3)$$

where V_n is the volume of the unit ball in \mathbb{R}^n .

Proof Let

$$g = \nabla Q(x_0), \text{ and } G = \nabla^2 Q. \quad (3.4)$$

Then

$$\begin{aligned} & |Q|_{H^1(\mathcal{B})}^2 \\ &= \int_{\|x-x_0\|_2 \leq r} \|G(x-x_0) + g\|_2^2 dx \\ &= \int_{\|x\|_2 \leq r} \|Gx + g\|_2^2 dx \\ &= \int_{\|x\|_2 \leq r} (x^T G^2 x + 2x^T Gg + \|g\|_2^2) dx. \end{aligned} \quad (3.5)$$

Because of symmetry, the integral of $x^T Gg$ is zero. Besides,

$$\int_{\|x\|_2 \leq r} \|g\|_2^2 dx = V_n r^n \|g\|_2^2. \quad (3.6)$$

Thus we only need to find the integral of $x^T G^2 x$. Without loss of generality, assume that G is diagonal (if not, apply a rotation). Denote the i -th diagonal entry of G by $G_{(ii)}$, and the i -th coordinate of x by $x_{(i)}$. Then

$$\int_{\|x\|_2 \leq r} x^T G^2 x dx = \int_{\|x\|_2 \leq r} \left(\sum_{i=1}^n G_{(ii)}^2 x_{(i)}^2 \right) dx = \|G\|_F^2 \int_{\|x\|_2 \leq r} x_{(1)}^2 dx. \quad (3.7)$$

To finish the proof, we show that

$$\int_{\|x\|_2 \leq r} x_{(1)}^2 dx = \frac{V_n r^{n+2}}{n+2}. \quad (3.8)$$

It suffices to justify (3.8) for the case $r = 1$ and $n \geq 2$. First,

$$\int_{\|x\|_2 \leq 1} x_{(1)}^2 dx = V_{n-1} \int_{-1}^1 u^2 (1-u^2)^{\frac{n-1}{2}} du = V_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2 \theta \cos^n \theta d\theta, \quad (3.9)$$

and, similarly,

$$V_n = V_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n \theta d\theta. \quad (3.10)$$

Second, integration by parts shows that

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n+2} \theta d\theta = (n+1) \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2 \theta \cos^n \theta d\theta. \quad (3.11)$$

Now it is easy to obtain (3.8) from (3.9-3.11). \square

Theorem 3.1 tells us that the H^1 seminorm of a quadratic function Q over a ball \mathcal{B} is closely related to a combination of $\|\nabla^2 Q\|_F^2$ and $\|\nabla Q(x_0)\|_2^2$, x_0 being the center of \mathcal{B} , and the combination coefficients being determined by the radius of \mathcal{B} . This result is interesting for two reasons. First, it enables us to measure the overall magnitude of the gradient over a ball. This is nontrivial for a quadratic function, since its gradient is not constant. Second, it clarifies the analytical and geometrical meaning of combining the Frobenius norm of the Hessian and the 2-norm of the gradient, and enables us to select the combination coefficients according to the geometrical meaning.

3.2 New Insights into Least-Norm Interpolation

The H^1 seminorm provides a new angle of view to understand least-norm interpolation. For convenience, we rewrite the least-norm interpolation problem here and call it the problem $P_1(\sigma)$:

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla^2 Q\|_F^2 + \sigma \|\nabla Q(x_0)\|_2^2 \\ \text{s.t.} \quad & Q(x) = f(x), \quad x \in \mathcal{S}. \end{aligned} \quad (P_1(\sigma))$$

We assume that the interpolation system

$$Q(x) = f(x), \quad x \in \mathcal{S} \quad (3.12)$$

is consistent on \mathcal{Q} , namely

$$\{Q \in \mathcal{Q} : Q(x) = f(x) \text{ for every } x \in \mathcal{S}\} \neq \emptyset. \quad (3.13)$$

With the help of Theorem 3.1, we see that the problem $P_1(\sigma)$ is equivalent to the problem

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & |Q|_{H^1(B_r)} \\ \text{s.t.} \quad & Q(x) = f(x), \quad x \in \mathcal{S} \end{aligned} \quad (P_2(r))$$

in some sense. The purpose of this subsection is to clarify the equivalence.

When $\sigma > 0$, the equivalence is easy and we state it as follows.

Theorem 3.2 *If $\sigma > 0$, then the least-norm interpolation problem $P_1(\sigma)$ is equivalent to the problem $P_2(r)$ with $r = \sqrt{(n+2)/\sigma}$.*

It turns out that the least-norm interpolation problem with positive σ essentially seeks the interpolant with minimal H^1 seminorm over a ball. The geometrical meaning of x_0 and σ is clear now: x_0 is the center of the ball, and $\sqrt{(n+2)/\sigma}$ is the radius.

Now we consider the least-norm interpolation problem with $\sigma = 0$. This case is particularly interesting, because it appears in several practical algorithms [14, 44, 50]. Since $P_1(\sigma)$ may have more than one solutions when $\sigma = 0$, we redefine $P_1(0)$ to be the bilevel least-norm interpolation problem

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla Q(x_0)\|_2 \\ \text{s.t.} \quad & \min_{Q \in \mathcal{Q}} \|\nabla^2 Q\|_F \\ \text{s.t.} \quad & Q(x) = f(x), \quad x \in \mathcal{S}. \end{aligned} \quad (P_1(0))$$

The redefinition is reasonable, because we have the following proposition.

Proposition 3.1 *For each $\sigma \geq 0$, the problem $P_1(\sigma)$ has a unique solution Q_σ . Moreover, Q_σ converges² to Q_0 when σ tends to 0^+ .*

² We define the convergence on \mathcal{Q} to be the convergence of coefficients.

Proof First, we consider the uniqueness of solution when $\sigma > 0$. For simplicity, denote

$$|Q|_\sigma = [\|\nabla^2 Q\|_F^2 + \sigma \|\nabla Q(x_0)\|_2^2]^{1/2}. \quad (3.14)$$

Let q_1 and q_2 be solutions of the problem $P_1(\sigma)$. Then the average of them is also a solution. According to the identity

$$\frac{1}{2}(|q_1|_\sigma^2 + |q_2|_\sigma^2) - \frac{1}{2}|(q_1 + q_2)|_\sigma^2 = \frac{1}{4}|q_1 - q_2|_\sigma^2, \quad (3.15)$$

we conclude that $|q_1 - q_2|_\sigma = 0$. Thus $\nabla^2 q_1 = \nabla^2 q_2$ and $\nabla q(x_0) = \nabla q_2(x_0)$. Now with the help of any one of the interpolation constraints $q_1(x) = f(x) = q_2(x)$, $x \in S$, we find that the quadratic functions q_1 and q_2 are identical as required.

When $\sigma = 0$, we can prove the uniqueness by similar argument, noticing the identity

$$\frac{1}{2}(\|\nabla^2 q_1\|_F^2 + \|\nabla^2 q_2\|_F^2) - \frac{1}{2}\|\nabla^2(q_1 + q_2)\|_F^2 = \frac{1}{4}\|\nabla^2 q_1 - \nabla^2 q_2\|_F^2, \quad (3.16)$$

and a similar one for the gradients.

Now we show that Q_σ converges to Q_0 when σ tends to 0^+ . Assume that the convergence does not hold, then we can pick a sequence of positive numbers $\{\sigma_k\}$ such that $\{\sigma_k\}$ tends to 0 but $\{Q_{\sigma_k}\}$ does not converge to Q_0 . Notice that

$$\|\nabla^2 Q_{\sigma_k}\|_F \geq \|\nabla^2 Q_0\|_F, \quad (3.17)$$

and

$$\|\nabla^2 Q_{\sigma_k}\|_F^2 + \sigma_k \|\nabla Q_{\sigma_k}(x_0)\|_2^2 \leq \|\nabla^2 Q_0\|_F^2 + \sigma_k \|\nabla Q_0(x_0)\|_2^2. \quad (3.18)$$

It follows that

$$\|\nabla Q_{\sigma_k}(x_0)\|_2 \leq \|\nabla Q_0(x_0)\|_2. \quad (3.19)$$

By (3.18–3.19) and the interpolation constraints, $\{Q_{\sigma_k}\}$ is bounded. Thus we can assume that $\{Q_{\sigma_k}\}$ converges to some \bar{Q} that is different from Q_0 . Then inequalities (3.17–3.18) implies $\|\nabla^2 \bar{Q}\|_F = \|\nabla^2 Q_0\|_F$, and (3.19) implies $\|\nabla \bar{Q}(x_0)\|_2 \leq \|\nabla Q_0(x_0)\|_2$. Hence \bar{Q} is another solution of the problem $P_1(0)$, contradicting to the uniqueness proved above. \square

Note that Proposition 3.1 implies the problem $P_2(r)$ has a unique solution for each positive r . Now we can state the relation between the problems $P_1(0)$ and $P_2(r)$.

Theorem 3.3 *When r tends to infinity, the solution of $P_2(r)$ converges to the solution of $P_1(0)$.*

Theorem 3.3 indicates that, when $\sigma = 0$, the least-norm interpolation problem seeks the interpolant with minimal H^1 seminorm over \mathbb{R}^n in the sense of limit. Thus Theorem 3.3 is the extension of Theorem 3.2 to the case $\sigma = 0$.

The questions listed in Section 1 can be answered now. The meaning of the objective function in least-norm interpolation is to minimize the H^1 seminorm of the interpolant over a ball. The reason to combine the Frobenius norm of the Hessian and the 2-norm of the gradient is to measure the H^1 seminorm of the interpolant. The parameters x_0 and σ determines the ball where the H^1 seminorm is calculated, x_0 being the center and $\sqrt{(n+2)/\sigma}$ being the radius. When $\sigma = 0$, we can interpret least-norm interpolation in the sense of limit.

4 On the Extended Symmetric Broyden Update

In this section, we employ the H^1 seminorm to study the extended symmetric Broyden update proposed by Powell [47]. As introduced in Section 2, the update defines Q_+ to be the solution of

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 + \sigma \|\nabla Q(x_0) - \nabla Q_0(x_0)\|_2^2 \\ \text{s.t.} \quad & Q(x) = F(x), \quad x \in \mathcal{S}, \end{aligned} \quad (4.1)$$

F being the objective function, and Q_0 being the model used in the previous trust-region iteration. When $\sigma = 0$, it is the symmetric Broyden update studied by Powell [43, 44, 45]. We focus on the case $\sigma > 0$. In Subsection 4.1, we interpret the update with the H^1 seminorm. In Subsection 4.2, we discuss the choices of x_0 and σ in the update, and show how to choose x_0 and σ according to their geometrical meaning. In Subsection 4.3, we test our choices of x_0 and σ through numerical experiments. It is worth mentioning that, the numerical experiments in Subsection 4.3 are designed in a special way in order to reduce the influence of computer rounding errors and obtain more convincing results.

4.1 Interpret the Update with the H^1 Seminorm

According to Theorem 3.2, problem (4.1) is equivalent to

$$\begin{aligned} \min_{Q \in \mathcal{Q}} \quad & |Q - Q_0|_{H^1(\mathcal{B})} \\ \text{s.t.} \quad & Q(x) = F(x), \quad x \in \mathcal{S}, \end{aligned} \quad (4.2)$$

where

$$\mathcal{B} = \left\{ x \in \mathbb{R}^n : \|x - x_0\|_2 \leq \sqrt{(n+2)/\sigma} \right\}. \quad (4.3)$$

Thus the extended symmetric Broyden update seeks the closest quadratic model to Q_0 subject to the interpolation constraints, distance being measured by the seminorm $|\cdot|_{H^1(\mathcal{B})}$.

Similar to the symmetric Broyden update, the extended symmetric Broyden update enjoys a very good approximation property when F is a quadratic function, as stated in Proposition 4.1.

Proposition 4.1 *If F is a quadratic function, then the solution Q_+ of problem (4.1) satisfies*

$$|Q_+ - F|_{H^1(\mathcal{B})}^2 = |Q_0 - F|_{H^1(\mathcal{B})}^2 - |Q_+ - Q_0|_{H^1(\mathcal{B})}^2, \quad (4.4)$$

where \mathcal{B} is defined as (4.3).

Proof Let $Q_t = Q_+ + t(Q_+ - F)$, t being any real number. Then Q_t is a quadratic function interpolating F on \mathcal{S} , as F is quadratic and Q_+ interpolates it. The optimality of Q_+ implies that the function

$$\varphi(t) = |Q_t - Q_0|_{H^1(\mathcal{B})}^2 \quad (4.5)$$

attains its minimum when t is zero. Expanding $\varphi(t)$, we obtain

$$\varphi(t) = t^2 |Q_+ - F|_{H^1(\mathcal{B})}^2 + 2t \int_{\mathcal{B}} [\nabla(Q_+ - Q_0)(x)]^T [\nabla(F - Q_0)(x)] dx + |Q_+ - Q_0|_{H^1(\mathcal{B})}^2. \quad (4.6)$$

Hence

$$\int_{\mathcal{B}} [\nabla(Q_+ - Q_0)(x)]^T [\nabla(F - Q_0)(x)] dx = 0. \quad (4.7)$$

Then we obtain (4.4) by considering $\varphi(-1)$. \square

In light of Theorem 3.1, equation (4.4) is equivalent to equation (1.9) of Powell [47]. Notice that equation (4.4) implies

$$\int_{\mathcal{B}} \|\nabla Q_+(x) - \nabla F(x)\|_2^2 dx \leq \int_{\mathcal{B}} \|\nabla Q_0(x) - \nabla F(x)\|_2^2 dx. \quad (4.8)$$

In other words, ∇Q_+ approximates ∇F on \mathcal{B} better than ∇Q_0 unless $\nabla Q_+ = \nabla Q_0$.

4.2 Choices of x_0 and σ

Now we turn our attention to the choices of x_0 and σ in the update. Recall that the purpose of the update is to construct a model for a trust-region subproblem. As in classical trust-region methods, the trust region is available before the update is applied, and we suppose it is

$$\{x : \|x - \bar{x}\|_2 \leq \Delta\}, \quad (4.9)$$

the point \bar{x} being the trust-region center, and the positive number Δ being the trust-region radius.

Powell [47] chooses x_0 and σ by exploiting the Lagrange functions of the interpolation problem (4.1). Suppose that $S = \{y_0, y_1, \dots, y_m\}$, then the i -th ($i = 0, 1, \dots, m$) Lagrange function of problem (4.1) is defined to be the solution of

$$\begin{aligned} \min_{l_i \in \mathcal{Q}} \quad & \|\nabla^2 l_i\|_F^2 + \sigma \|\nabla l_i\|_2^2 \\ \text{s.t. } \quad & l_i(y_j) = \delta_{ij}, \quad j = 0, 1, \dots, m, \end{aligned} \quad (4.10)$$

δ_{ij} being the Kronecker delta. In the algorithm of Powell [47], S is maintained in a way so that Q_0 interpolates F on S except one point, say y_0 without loss of generality. Then the interpolation constraints can be rewritten as

$$Q(y_j) - Q_0(y_j) = [F(y_0) - Q_0(y_0)] \delta_{0j}, \quad j = 0, 1, \dots, m. \quad (4.11)$$

Therefore the solution of problem (4.1) is

$$Q_+ = Q_0 + [F(y_0) - Q_0(y_0)] l_0. \quad (4.12)$$

Hence l_0 plays a central part in the update. Thus Powell [47] chooses x_0 and σ by examining l_0 . Powell shows that, in order to make l_0 behaves well, the ratio $\|x_0 - \bar{x}\|_2 / \Delta$ should not be much larger than one, therefore x_0 is set to be \bar{x} throughout the calculation. The choice of σ is a bit complicated. The basic idea is to balance $\|\nabla^2 l_0\|_F^2$ and $\sigma \|\nabla l_0(x_0)\|_2^2$. Thus σ is set to η / ξ , where η and ξ are estimates of the magnitudes of $\|\nabla^2 l_0\|_F^2$ and $\|\nabla l_0(x_0)\|_2^2$, respectively. See Powell [47] for details.

In Subsection 4.1, we have interpreted the extended symmetric Broyden update with the H^1 seminorm. The new interpretation enables us to choose x_0 and σ in a geometrical way. According to problem (4.2), choosing x_0 and σ is equivalent to choosing the ball \mathcal{B} . Let us think about the role that \mathcal{B} plays in the update. First, \mathcal{B} is the region where the update tries to preserve information from Q_0 , by minimizing the change with respect to the seminorm

$|\cdot|_{H^1(\mathcal{B})}$. Second, \mathcal{B} is the region where the update tends to improve the model, as suggested by the facts (4.4) and (4.8) for quadratic objective function. Thus we should choose \mathcal{B} to be a region where the behavior of the new model is important to us. It may seem adequate to pick \mathcal{B} to be the trust region (4.9). But we prefer to set it bigger, because the new model Q_+ will influence its successors via the least change update, and thereby influence subsequent trust-region iterations. Thus it is myopic to consider only the current trust region, and a more sensible choice is to let \mathcal{B} be the ball $\{x : \|x - \bar{x}\|_2 \leq M\Delta\}$, M being a positive number bigger than one. Moreover, we find in practice that it is helpful to require $\mathcal{B} \supset \mathcal{S}$. Thus our choice of \mathcal{B} is

$$\{x : \|x - \bar{x}\|_2 \leq r\}, \quad (4.13)$$

where

$$r = \max \left\{ M\Delta, \max_{x \in \mathcal{S}} \|x - \bar{x}\|_2 \right\}. \quad (4.14)$$

Consequently, we set

$$x_0 = \bar{x}, \text{ and } \sigma = \frac{n+2}{r^2}. \quad (4.15)$$

Therefore our choice of x_0 coincides with Powell's, but the choice of σ is different.

4.3 Numerical Results

With the help of the H^1 seminorm, we have proposed a new method of choosing σ for the extended symmetric Broyden update. In this subsection we test the new σ through numerical experiments, and make comparison with the one by Powell [47]. We will also observe whether positive σ in (4.1) brings improvements or not.

Powell [47] implements a trust-region algorithm based on the extended symmetric Broyden update in Fortran 77 (it is an extended version of the software BOBYQA [46]), and we use this code for our experiments. We modify this code to get three solvers as follows for comparison.

- a.) SYMB: σ is always set to zero. It is equivalent to using the original symmetric Broyden update.
- b.) ESYMBP: σ is chosen according to Powell [47], as described in the second paragraph of Subsection 4.2.
- c.) ESYMBS: σ is chosen according to (4.14–4.15). We set $M = 10$ in (4.14).

In the code of Powell [47], the size of the interpolation set \mathcal{S} keeps unchanged throughout the computation, and the user can set it to be any integer between $n+2$ and $(n+1)(n+2)/2$. We choose $2n+1$ as recommended. The code of Powell [47] is capable of solving problems with box constraints. We set the bounds to infinity since we are focusing on unconstrained problems. In this case, SYMB can be regarded as another implementation of NEWUOA [44].

We assume that evaluating the objective function is the most expensive part for optimization without derivatives. Thus we might compare the performance of different derivative-free solvers by simply counting the numbers of function evaluations they use until termination, provided that they have found the same minimizer to high accuracy. However, this is inadequate in practice, because, according to our numerical experiences, computer rounding errors could substantially influence the number of function evaluations. Here we present an

interesting example, which is inspired by the numerical experiments of Powell [47]. Consider the test problem BDQRTIC [9] with the objective function

$$F(x) = \sum_{i=1}^{n-4} [(x_i^2 + 2x_{i+1}^2 + 3x_{i+2}^2 + 4x_{i+3}^2 + 5x_n^2)^2 - 4x_i + 3], \quad x \in \mathbb{R}^n, \quad (4.16)$$

and the starting point $e = (1 \ 1 \ \dots \ 1)^T$. Let P be a permutation matrix on \mathbb{R}^n , and define $\tilde{F}(x) = F(Px)$. In theory, the solver SYMB should perform equally on F and \tilde{F} with initial vectors both set to e , as the implementation we are using is mathematically independent of variable ordering. But it is not the case in practice. For $n = 20$, we applied SYMB to F and \tilde{F} with P chosen specifically³. The solver terminated with function values 3.540906874607717D+01 and 3.540906874698564D+01, but the numbers of function evaluations were 760 and 4267, respectively. To make things more interesting, we applied ESYMBs to the same problems. This time the function values were 3.540906874638613D+01 and 3.540906874609878D+01, while the numbers of function evaluations were 4124 and 833. Notice that this dramatic story was entirely a monodrama of computer rounding errors. Similar story also happens to other derivative-free solvers. For instance, we permuted the variables in BDQRTIC ($n = 20$) with the method described above, and then solved the permuted variants of BDQRTIC with the build-in function `fminsearch` of MATLAB 7.6.0.324 (R2008a), which is a derivative-free solver based on the Nelder-Mead simplex direct search [35] described in Lagarias, Reeds, Wright, and Wright [32]. The numbers of function evaluations varied from 9217 to 70105 during the experiments⁴. This example might reflect a common difficulty for derivative-free solvers. It suggests we pay enough attention to the instability caused by computer rounding errors when designing numerical tests for solvers of this type.

We test the solvers SYMB, ESYMBs, and ESYMBP with the following method, in order to observe their performance in a relatively reliable way, and to inspect their stability to some extent. The basic idea is from the numerical experiments of Powell [47]. Given a test problem \mathcal{P} with objective function F and starting point \hat{x} , we randomly generate N permutation matrices P_i ($i = 1, 2, \dots, N$), and let

$$F_i(x) = F(P_i x), \quad \hat{x}_i = P_i^{-1} \hat{x}, \quad i = 1, 2, \dots, N. \quad (4.17)$$

Then we employ the solvers to minimize F_i starting from \hat{x}_i . For solver S , we obtain a vector $\#F = (\#F_1, \#F_2, \dots, \#F_N)$, $\#F_i$ being the number of function evaluations required by S for F_i and \hat{x}_i . If S fails on F_i and \hat{x}_i , we replace $\#F_i$ with $K\#F_i$ as a penalty, K being a number bigger than one. Then compute the mean value and the standard deviation of vector $\#F$:

$$\text{mean}(\#F) = \frac{1}{N} \sum_{i=1}^N \#F_i, \quad (4.18)$$

and

$$\text{std}(\#F) = \sqrt{\frac{1}{N} \sum_{i=1}^N [\#F_i - \text{mean}(\#F)]^2}. \quad (4.19)$$

³ The matrix P was $(e_8 \ e_5 \ e_{13} \ e_{20} \ e_4 \ e_{18} \ e_{10} \ e_1 \ e_7 \ e_{19} \ e_{12} \ e_9 \ e_{14} \ e_3 \ e_{17} \ e_6 \ e_{15} \ e_{16} \ e_{11} \ e_2)$, e_i being the i -th coordinate vector in \mathbb{R}^{20} .

⁴ The matrix P for 9217 was $(e_{15} \ e_7 \ e_5 \ e_{18} \ e_2 \ e_{19} \ e_6 \ e_{20} \ e_{13} \ e_4 \ e_8 \ e_1 \ e_9 \ e_{12} \ e_{10} \ e_{14} \ e_{17} \ e_{11} \ e_{16} \ e_3)$, and the one for 70105 was $(e_3 \ e_5 \ e_{15} \ e_9 \ e_{16} \ e_6 \ e_{18} \ e_{19} \ e_2 \ e_{10} \ e_4 \ e_{11} \ e_{13} \ e_{20} \ e_{17} \ e_1 \ e_7 \ e_8 \ e_{14} \ e_{12})$, e_i being the i -th coordinate vector in \mathbb{R}^{20} .

When N is reasonably large, $\text{mean}(\#F)$ estimates the average performance of solver S on problem \mathcal{P} , and $\text{std}(\#F)$ reflects the stability of solver S on problem \mathcal{P} . We may also compute the relative standard deviation of $\#F$, namely,

$$\text{rstd}(\#F) = \frac{\text{std}(\#F)}{\text{mean}(\#F)}, \quad (4.20)$$

to obtain a normalized measure of stability. Besides, $\min(\#F)$ and $\max(\#F)$ are meaningful. They approximate the best-case and worst-case performance of solver S on problem \mathcal{P} . By comparing these statistics, we can reasonably assess the solvers under consideration. We use $N = 10$ in practice.

With the method addressed above, we tested SYMB, ESYMBP, and ESYMBS on BDQRTIC and six test problems used by Powell [44], namely, ARWHEAD, CHROSEN, PENALTY1, PENALTY2, SPHRPTS, and VARDIM. The starting points and the initial/final trust-region radius were set according to Powell [42, 44]. See Powell [42, 44] for details of these test problems. For each problem, we solved it for dimensions 10, 12, 14, 16, 18, and 20. The experiments were carried out on a Dell PC with Linux, and the compiler was gfortran in GCC 4.3.3.

All the solvers terminated successfully for every problem with every dimension, except the problem CHROSEN. The objective function of CHROSEN is the “chained Rosenbrock” function

$$F(x) = \sum_{i=1}^{n-1} [4(x_i - x_{i+1}^2)^2 + (1 - x_{i+1})^2], \quad x \in \mathbb{R}^n, \quad (4.21)$$

and the starting point is $\hat{x} = -(1 \ 1 \ \cdots \ 1)^T$. Its global minimizer is $e = (1 \ 1 \ \cdots \ 1)$, which is difficult to find. When a solver found a local minimizer different from e , we called it a failure and penalized⁵ the number of function evaluations by multiplying 3. In Table 4.1, the number corresponding to a dimension n and a solver S presents how many times S failed when solving the ten permuted variants of n -dimensional CHROSEN. All the solvers failed for several times.

Table 4.1: Number of Failures on CHROSEN

n	10	12	14	16	18	20
SYMB	3	1	1	2	4	4
ESYMBP	0	3	3	0	2	4
ESYMBS	0	1	1	1	0	4

Table 4.2 reports the values of $\text{mean}(\#F)$ generated in our experiments. For each problem and each dimension, we present the statistics in the order SYMB (the first line), ESYMBP (the second line), and ESYMBS (the third line). The values have been rounded to the nearest integer. Let us first focus on the comparison between ESYMBP and ESYMBS. Table 4.2 suggests that ESYMBS worked better than ESYMBP in the sense of average performance. Thus our choice of σ improved the performance of the extended symmetric Broyden update on the test problems. It convinces us that the H^1 seminorm does bring a better interpretation

⁵ Although the solvers under consideration are not designed for global optimization, we still think the penalization is necessary, because when different minimizers are found, the corresponding numbers of function evaluations are not comparable.

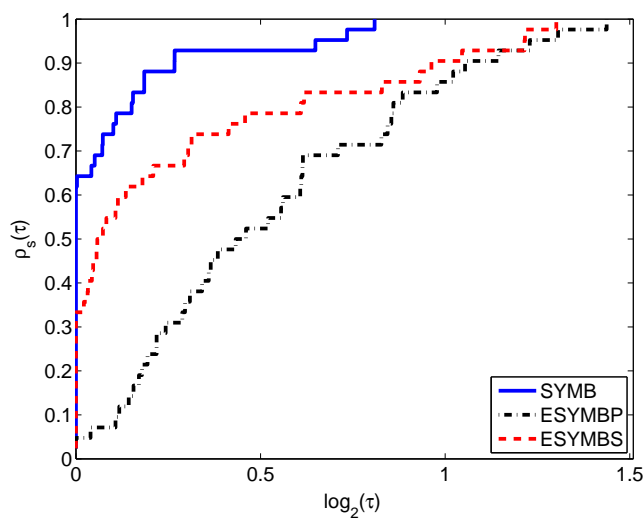
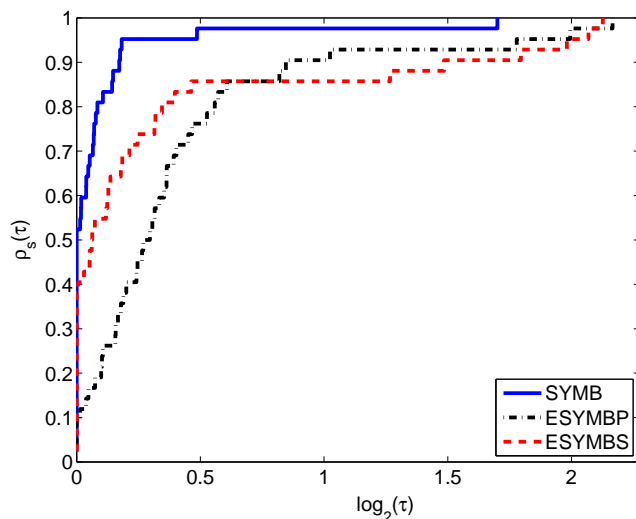
of the update. Compared with SYMB, both ESYMBP and ESYMBS usually worked worse in the sense of average performance. Thus both choices of positive σ rarely reduced the number of function evaluations in our experiments. However, we should notice that the positive σ did not do much damage, even though very large σ was allowed (for example, the σ in ESYMBS could be as large as 10^{12} in our experiments).

Table 4.2: $\text{mean}(\#F)$ of SYMB, ESYMBP, and ESYMBS

n	10	12	14	16	18	20
ARWHEAD	186	239	338	426	435	784
	274	431	393	461	552	806
	231	328	417	425	503	796
BDQRTIC	1017	1087	1748	1381	1266	1794
	1399	1929	2406	2804	2967	3221
	916	1659	2143	2692	2414	3187
CHROSEN	713	679	819	1113	1753	1861
	515	1193	1340	925	1471	2165
	455	646	862	962	1001	2013
PENALTY1	2900	3599	4360	5637	6265	7752
	4419	4698	5886	6012	7689	8507
	3019	3877	4458	5255	6471	7188
PENALTY2	588	1656	1018	1515	1476	1479
	1299	1609	2516	2320	3065	4006
	1213	1457	2365	2325	3427	3644
SPHRPTS	250	397	3285	1831	1598	2708
	492	607	3503	1997	1999	2382
	256	436	2730	1100	1552	3173
VARDIM	1550	2399	3371	3944	4615	5873
	1997	2828	3611	4394	5716	6327
	1392	2316	3208	4064	5230	6219

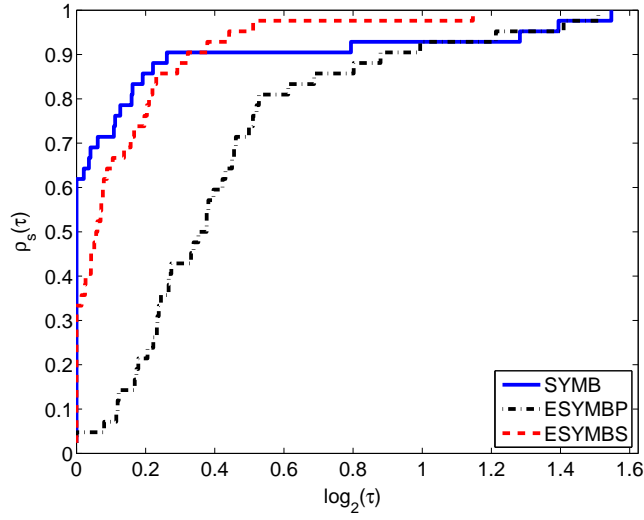
We visualize the comparison made above by presenting the performance profile [23] of $\text{mean}(\#F)$ in Fig. 4.1. The same as traditional performance profiles, the higher means the better. The performance profile supports the conclusions made above. Additionally, we give the performance profiles of $\text{min}(\#F)$ and $\text{max}(\#F)$ in Fig. 4.2 and Fig. 4.3, respectively. We notice that ESYMBP and ESYMBS hold better positions in Fig. 4.2 than in Fig. 4.1. Since $\text{min}(\#F)$ reflects the best-case performance of solvers, we infer that, if all the solvers could perform their best, then both ESYMBP and ESYMBS would become more competitive. It suggests that these two solvers still have the potential to improve.

Table 4.3 shows the values of $\text{std}(\#F)$ and $\text{rstd}(\#F)$ generated in our experiments. We present the statistics in the same order as in Table 4.2. For each solver and each problem, the values of $\text{std}(\#F)$ and $\text{rstd}(\#F)$ come together in the form “ $\text{std}(\#F)/\text{rstd}(\#F)$ ”. The values of $\text{std}(\#F)$ are rounded to the nearest integer, and only three digits of $\text{rstd}(\#F)$ are shown. It is clear that all the three solvers often suffered from computer rounding errors, except when solving ARWHEAD. Notice that the big values of $\text{std}(\#F)$ and $\text{rstd}(\#F)$ for CHROSEN are partially due to the penalization on failures (see Table 4.1). With the help of the performance profiles of $\text{std}(\#F)$ and $\text{rstd}(\#F)$ presented in Fig. 4.4 and Fig. 4.5, we conclude that ESYMBS performed relatively stabler than its competitors on the test problems.

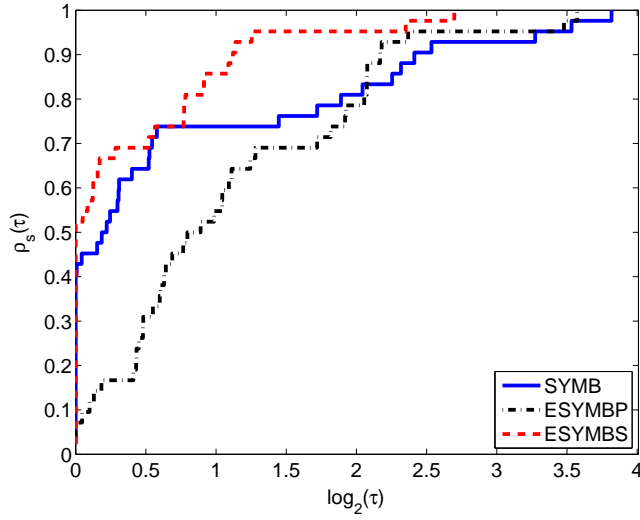
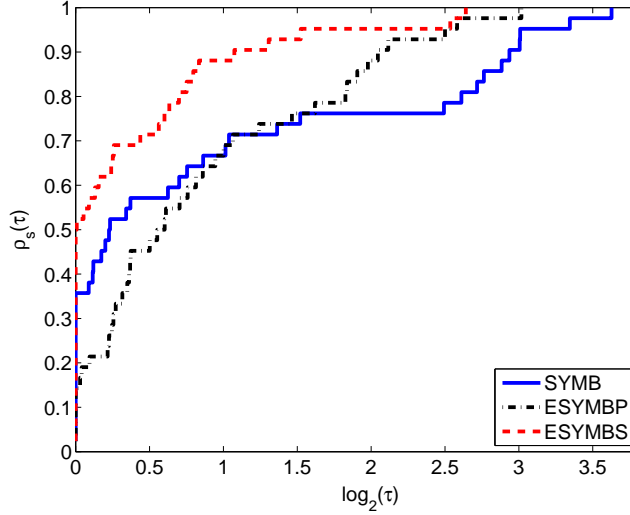
Fig. 4.1: Performance Profile of $\text{mean}(\#F)$ Fig. 4.2: Performance Profile of $\text{min}(\#F)$

5 Conclusions

Least-norm interpolation has wide applications in derivative-free optimization. Motivated by its objective function, we have studied the H^1 Sobolev seminorm of quadratic functions. We give an explicit formula for the H^1 seminorm of quadratic functions over balls. It turns out that least-norm interpolation essentially seeks a quadratic interpolant with minimal H^1 seminorm over a ball. Moreover, we find that the parameters x_0 and σ in the objective

Fig. 4.3: Performance Profile of $\max(\#F)$ Table 4.3: $\text{std}(\#F)$ and $\text{rstd}(\#F)$ of SYMB, ESYMBP, and ESYMBS

n	10	12	14	16	18	20
ARWHEAD	4/0.02	7/0.03	9/0.03	24/0.06	36/0.08	24/0.03
	18/0.07	27/0.06	36/0.09	32/0.07	46/0.08	58/0.07
	8/0.03	17/0.05	19/0.05	25/0.06	25/0.05	41/0.05
BDQRTIC	384/0.38	676/0.62	713/0.41	897/0.65	762/0.60	1355/0.76
	257/0.18	182/0.09	602/0.25	728/0.26	1158/0.39	1230/0.38
	485/0.53	388/0.23	143/0.07	680/0.25	1127/0.47	1191/0.37
CHROSEN	458/0.64	444/0.65	474/0.58	710/0.64	1048/0.60	1102/0.59
	54/0.11	775/0.65	844/0.63	73/0.08	884/0.60	1206/0.56
	40/0.09	359/0.56	426/0.49	476/0.50	74/0.07	1228/0.61
PENALTY1	259/0.09	258/0.07	344/0.08	489/0.09	338/0.05	467/0.06
	349/0.08	547/0.12	709/0.12	557/0.09	496/0.06	893/0.10
	282/0.09	257/0.07	374/0.08	419/0.08	735/0.11	377/0.05
PENALTY2	326/0.55	101/0.06	697/0.68	1006/0.66	1059/0.72	1014/0.69
	454/0.35	426/0.26	240/0.10	1286/0.55	701/0.23	396/0.10
	119/0.10	516/0.35	211/0.09	849/0.37	199/0.06	246/0.07
SPHRPTS	14/0.06	56/0.14	410/0.12	354/0.19	122/0.08	1230/0.45
	155/0.31	237/0.39	712/0.20	245/0.12	630/0.32	954/0.40
	17/0.07	96/0.22	461/0.17	74/0.07	175/0.11	212/0.07
VARDIM	225/0.14	270/0.11	316/0.09	478/0.12	481/0.10	447/0.08
	265/0.13	409/0.14	652/0.18	543/0.12	520/0.09	623/0.10
	156/0.11	262/0.11	335/0.10	389/0.10	336/0.06	768/0.12

Fig. 4.4: Performance Profile of $\text{std}(\#F)$ Fig. 4.5: Performance Profile of $\text{rstd}(\#F)$

function determines the center and radius of the ball, respectively. These insights further our understanding of least-norm interpolation.

With the help of the new observation, we have studied the extended symmetric Broyden update (4.1) proposed by Powell [47]. Since the update calculates the change to the old model by a least-norm interpolation, we can interpret it with our new theory. We have discussed how to choose x_0 and σ for the update according to their geometrical meaning. This discussion leads to the same x_0 used by Powell [47], and a very easy way to choose σ .

According to our numerical results, the new σ works relatively better than the one proposed by Powell [47].

It is interesting to ask whether the extended symmetric Broyden update can bring improvements to the original version (2.8) in derivative-free optimization. Until now, we can not give a positive answer. However, the numerical results of Powell [47] suggest that the extended version helps to improve the convergence rate. It would be desirable to theoretically investigate the local convergence properties of the updates. Results like (2.10) and (4.4) may help in such research. We think it is still too early to draw any conclusion on the extended symmetric Broyden update, unless we could find strong theoretical evidence.

The numerical example on BDQRTIC deserves serious attention. It shows clearly that we should be very careful when comparing derivative-free solvers via numerical experiments. Our method of designing numerical tests seems reasonable. It reduces the influence of computer rounding errors, and meanwhile reflects the stability of solvers. We hope this method can lead to more reliable benchmarking of derivative-free solvers.

Acknowledgements The author thanks Professor M. J. D. Powell for providing references and codes, for kindly answering the author's questions about the excellent software NEWUOA, and for the helpful discussions. Professor Powell helped improve the proofs of Theorem 3.1 and Proposition 3.1. The author owes much to his supervisor Professor Ya-Xiang Yuan for the valuable directions and suggestions. This work was partly finished during a visit to Professor Klaus Schittkowski at Universität Bayreuth in 2010, where the author has received pleasant support for research. The author thanks Humboldt Foundation for the financial assistance during this visit, and he is much more than grateful to Professor Schittkowski for his warm hospitality.

References

1. Adams, R.A.: Sobolev spaces. Pure and applied mathematics. Academic Press (1975)
2. Booker, A.J., Dennis, J.E., Frank, P.D., Serafini, D.B., Torczon, V., Trosset, M.W.: A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization* **17**(1), 1–13 (1999)
3. Booker, A.J., Dennis Jr, J.E., Frank, P., Serafini, D.B., Torczon, V., Trosset, M.W.: Optimization using surrogate objectives on a helicopter test example. *PROGRESS IN SYSTEMS AND CONTROL THEORY* **24**, 49–58 (1998)
4. Booker, A.J., Frank, P., Dennis Jr, J.E., Moore, D.W., Serafini, D.B.: Managing surrogate objectives to optimize helicopter rotor design—further experiments. In: *IN PROCEEDINGS OF THE SEVENTH AIAA/USAF/NASA/ISSMO SYMPOSIUM ON MULTIDISCIPLINARY ANALYSIS AND OPTIMIZATION* (1998)
5. Bortz, D., Kelley, C.T.: The simplex gradient and noisy optimization problems. *Progress in Systems and Control Theory* **24**, 77–90 (1998)
6. Brent, R.P.: Algorithms for minimization without derivatives. Prentice-Hall, Englewood Cliffs, New Jersey (1973)
7. Broyden, C.G., Dennis Jr, J.E., Moré, J.J.: On the local and superlinear convergence of quasi-newton methods. *IMA Journal of Applied Mathematics* **12**(3), 223–245 (1973)
8. Choi, T.D., Kelley, C.T.: Superlinear convergence and implicit filtering. *SIAM Journal on Optimization* **10**(4), 1149–1162 (2000)
9. Conn, A.R., Gould, N.I.M., Lescrenier, M., Toint, P.L.: Performance of a multifrontal scheme for partially separable optimization. In: S. Gomez, J.P. Hennart, R.A. Tapia (eds.) *Advances in optimization and numerical analysis, Proceedings of the Sixth workshop on Optimization and Numerical Analysis*, Oaxaca, Mexico, 1992, *Mathematics and its Applications Series*, vol. 275, pp. 79–96. Kluwer Academic, Dordrecht, Netherlands (1994)
10. Conn, A.R., Gould, N.I.M., Toint, P.L.: LANCELOT: a Fortran package for large-scale nonlinear optimization (release A). Springer series in computational mathematics. Springer-Verlag (1992)
11. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-region methods, *MPS-SIAM series on optimization*, vol. 1. Society for Industrial Mathematics (2000)

12. Conn, A.R., Scheinberg, K., Toint, P.L.: On the Convergence of Derivative-Free Methods for Unconstrained Optimization. In: M.D. Buhmann, A. Iserles (eds.) *Approximation Theory and Optimization: Tributes to M. J. D. Powell*, pp. 83–108. Cambridge University Press, Cambridge, United Kingdom (1997)
13. Conn, A.R., Scheinberg, K., Toint, P.L.: Recent progress in unconstrained nonlinear optimization without derivatives. *Mathematical Programming* **79**, 397–414 (1997)
14. Conn, A.R., Scheinberg, K., Toint, P.L.: A derivative free optimization algorithm in practice. In: *Proceedings of 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. St. Louis, MO (1998)
15. Conn, A.R., Scheinberg, K., Vicente, L.N.: *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2009)
16. Conn, A.R., Toint, P.L.: An Algorithm Using Quadratic Interpolation for Unconstrained Derivative Free Optimization. In: G. Di Pillo, F. Giannessi (eds.) *Nonlinear Optimization and Applications*, pp. 27–47. Kluwer Academic/Plenum Publishers, New York (1996)
17. Custódio, A.L., Dennis Jr, J.E., Vicente, L.N.: Using simplex gradients of nonsmooth functions in direct search methods. *IMA journal of numerical analysis* **28**(4), 770–784 (2008)
18. Custódio, A.L., Rocha, H., Vicente, L.N.: Incorporating minimum frobenius norm models in direct search. *Computational Optimization and Applications* **46**(2), 265–278 (2010)
19. Custódio, A.L., Vicente, L.N.: SID-PSM: A pattern search method guided by simplex derivatives for use in derivative-free optimization (2005)
20. Custódio, A.L., Vicente, L.N.: Using sampling and simplex derivatives in pattern search methods. *SIAM Journal on Optimization* **18**(2), 537–555 (2007)
21. Dennis Jr, J.E., Schnabel, R.B.: Least change secant updates for quasi-newton methods. *Siam Review* pp. 443–459 (1979)
22. Diniz-Ehrhardt, M.A., Martínez, J.M., Raydán, M.: A derivative-free nonmonotone line-search technique for unconstrained optimization. *Journal of Computational and Applied Mathematics* **219**(2), 383–397 (2008)
23. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. *Mathematical Programming* **91**(2), 201–213 (2002)
24. Evans, L.C.: *Partial differential equations*. Graduate studies in mathematics. American Mathematical Society (1998)
25. Fowler, K.R., Reese, J.P., Kees, C.E., Dennis Jr, J.E., Kelley, C.T., Miller, C.T., Audet, C., Booker, A.J., Couture, G., Darwin, R.W.: Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. *Advances in Water Resources* **31**(5), 743–757 (2008)
26. Gilmore, P., Kelley, C.T.: An implicit filtering algorithm for optimization of functions with many local minima. *SIAM Journal on Optimization* **5**, 269 (1995)
27. Hemker, T., Fowler, K., Von Stryk, O.: Derivative-free optimization methods for handling fixed costs in optimal groundwater remediation design. In: *Proc. of the CMWR XVI-Computational Methods in Water Resources*, pp. 19–22 (2006)
28. Kelley, C.: *Iterative methods for optimization*, *Frontiers in applied mathematics*, vol. 18. Society for Industrial Mathematics (1999)
29. Kelley, C.T.: A brief introduction to implicit filtering. North Carolina State Univ., Raleigh, NC, CRSC Tech. Rep. CRSC-TR02-28 (2002)
30. Kelley, C.T.: *Implicit filtering*. Society for Industrial and Applied Mathematics (2011)
31. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: New perspectives on some classical and modern methods. *Siam Review* pp. 385–482 (2003)
32. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization* **9**(1), 112–147 (1998)
33. Lewis, R.M., Torczon, V., Trosset, M.W.: Direct search methods: then and now. *Journal of Computational and Applied Mathematics* **124**(1), 191–207 (2000)
34. Marazzi, M., Nocedal, J.: Wedge trust region methods for derivative free optimization. *Mathematical programming* **91**(2), 289–305 (2002)
35. Nelder, J.A., Mead, R.: A simplex method for function minimization. *The Computer Journal* **7**(4), 308–313 (1965). DOI 10.1093/comjnl/7.4.308
36. Oeuvray, R.: *Trust-region methods based on radial basis functions with application to biomedical imaging*. Ph.D. thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (2005)
37. Oeuvray, R., Bierlaire, M.: A new derivative-free algorithm for the medical image registration problem. *International Journal of Modelling and Simulation* **27**(2), 115–124 (2007)
38. Oeuvray, R., Bierlaire, M.: a derivative-free algorithm based on radial basis functions. In: *International Journal of Modelling and Simulation* (2008)

39. Powell, M.J.D.: A direct search optimization method that models the objective and constraint functions by linear interpolation. *Advances in Optimization and Numerical Analysis* pp. 51–67 (1994)
40. Powell, M.J.D.: Direct search algorithms for optimization calculations. *Acta Numerica* **7**(1), 287–336 (1998)
41. Powell, M.J.D.: UOBYQA: unconstrained optimization by quadratic approximation. Tech. Rep. DAMTP NA2000/14, CMS, University of Cambridge (2000)
42. Powell, M.J.D.: On trust region methods for unconstrained minimization without derivatives. *Mathematical programming* **97**(3), 605–623 (2003)
43. Powell, M.J.D.: Least frobenius norm updating of quadratic models that satisfy interpolation conditions. *Mathematical Programming* **100**, 183–215 (2004)
44. Powell, M.J.D.: The NEWUOA software for unconstrained optimization without derivatives. Tech. Rep. DAMTP NA2004/08, CMS, University of Cambridge (2004)
45. Powell, M.J.D.: Developments of NEWUOA for minimization without derivatives. *IMA Journal of Numerical Analysis* pp. 649–664 (2008). DOI 10.1093/imanum/drm047
46. Powell, M.J.D.: The BOBYQA algorithm for bound constrained optimization without derivatives. Tech. Rep. DAMTP 2009/NA06, CMS, University of Cambridge (2009)
47. Powell, M.J.D.: Beyond symmetric Broyden for updating quadratic models in minimization without derivatives. Tech. Rep. DAMTP 2010/NA04, CMS, University of Cambridge (2010)
48. Stewart III, G.: A modification of davidon’s minimization method to accept difference approximations of derivatives. *Journal of the ACM (JACM)* **14**(1), 72–83 (1967)
49. Vanden Berghen, F., Bersini, H.: CONDOR, a new parallel, constrained extension of powell’s UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics* **181**(1), 157–175 (2005)
50. Wild, S.M.: MNH: a derivative-free optimization algorithm using minimal norm Hessians. In: Tenth Copper Mountain Conference on Iterative Methods (2008)
51. Wild, S.M., Regis, R.G., Shoemaker, C.A.: ORBIT: Optimization by radial basis function interpolation in trust-regions. *SIAM Journal on Scientific Computing* **30**(6), 3197–3219 (2008)
52. Winfield, D.H.: Function minimization by interpolation in a data table. *IMA Journal of Applied Mathematics* **12**(3), 339 (1973)
53. Wright, M.H.: Direct search methods: Once scorned, now respectable. *PITMAN RESEARCH NOTES IN MATHEMATICS SERIES* pp. 191–208 (1996)